推定可能関数によるTypeII，TypeIII，TypeIVの違いの説明

三輪哲久
国立研究開発法人　農業環境技術研究所

TypeII, TypeIII and TypeIV Estimable Functions

Tetsuhisa Miwa
National Institute for Agro-Environmental Sciences

要旨

アンバランストな多元配置分散分析モデルに対して，PROC GLM（および他のいくつかのプロシージャ）では，4つのタイプ（TypeI，II，III，IV）の平方和が計算される。我国の多くの文献では，パラメータに対する制約条件の与え方によって異なるタイプの平方和が計算されると説明されている。そのため，これら4つのタイプの考え方に対して様々な混乱が生じている。しかし，これらのタイプの本質的な違いは，何を推定しているか（つまり，どのような推定可能関数を考えているか）によって生じているのである。本講演では，推定可能関数の観点からタイプの違いについて分かりやすく解説する。

キーワード: 線形モデル，制約条件，アンバランスト・モデル，多元配置，直交表

1　はじめに

アンバランストな多元配置モデルの分散分析においては，異なるタイプ (TypeI, II, III, IV) の平方和を考えることができる。しかし，その解釈に関しては次のような様々な意見があり，現在もなお混乱が続いている。

- TypeII の平方和では，最初に主効果だけを取り込んだモデルを考えている。
- 主効果については，TypeII の平方和の方が検出力が高い。
- 極端にアンバランストなときには，TypeII を使った方がよい。
- 欠測セルがあるときに，全てのセルを平等に扱う TypeIII は不自然である。
- TypeIII の平方和は特殊な制約条件に依存する。

多くの文献では，パラメータに対する制約条件や，モデル平方和の増加量を用いてタイプの違いが説明されている。特に TypeII と TypeIII について論争が行なわれている。

一方，SAS/STAT 9.3 ユーザーズガイド (2011) では，「Chapter 15: The Four Types of Estimable Functions (4つのタイプの推定可能関数)」において4つのタイプが説明されている。この章のタイトルが示すように，4つのタイプの違いは，本質的には線形モデルにおける推定可能関数の違いなのである。

2　線形モデルと推定可能関数

本節で線形モデルの基本事項を整理する（証明については，三輪 (2015) などを参照)。

2.1　線形モデルと正規方程式

線形モデル

$$\boldsymbol{y}_{N\times 1} = X_{N\times p}\boldsymbol{\theta}_{p\times 1} + \boldsymbol{e}_{N\times 1} \tag{1}$$

$\boldsymbol{y}$	$=$	$(y_1,\ldots,y_N)^T$	観測値ベクトル（N 次元）
X	$=$	$\{x_{ij}\}\quad (i=1,\ldots,N; j=1,\ldots,p)$	計画行列（既知定数）
$\boldsymbol{\theta}$	$=$	$(\theta_1,\ldots,\theta_p)^T$	未知パラメータ（p 次元）
$\boldsymbol{e}$	$=$	$(e_1,\ldots,e_N)^T \sim N(\mathbf{0},\ \sigma^2 I_N)$	観測誤差（正規分布）

最小二乗法

$$\sum_{i=1}^{N}\big(y_i-\sum_{j=1}^{p}x_{ij}\,\theta_j\big)^2=(\boldsymbol{y}-X\boldsymbol{\theta})^T(\boldsymbol{y}-X\boldsymbol{\theta})\implies \text{最小} \tag{2}$$

正規方程式

$$X^TX\hat{\boldsymbol{\theta}}=X^T\boldsymbol{y} \tag{3}$$

- X^TX が正則（X が列に関してフルランク）。
  $\hat{\boldsymbol{\theta}}=(X^TX)^{-1}X^T\boldsymbol{y}$
- X^TX が特異（X が列に関してフルランクではない，$\mathrm{rank}(X^TX)=\mathrm{rank}(X)=r<p$）。
  正規方程式は解をもつ（$\hat{\boldsymbol{\theta}}$ は一意には決まらない）。
  $\hat{\boldsymbol{\theta}}$ に対して $p-r$ 個の制約条件を付けることにより，$\hat{\boldsymbol{\theta}}$ を一応は一意的に決めることができる。

2.2 推定可能関数

パラメータ $\boldsymbol{\theta}$ の線形関数

$$\boldsymbol{l}^T\boldsymbol{\theta}=l_1\theta_1+\cdots+l_p\theta_p,\qquad \boldsymbol{l}^T=(l_1,\ldots,l_p) \tag{4}$$

で，$\boldsymbol{l}^T$ が X の横ベクトルの線形結合で表わされるもの，すなわち

$$\boldsymbol{l}^T=(l_1,\ldots,l_p)=(c_1,\ldots,c_N)\,X=\boldsymbol{c}^T\,X\quad(\boldsymbol{l}=X^T\,\boldsymbol{c}) \tag{5}$$

を満たす $\boldsymbol{l}^T\boldsymbol{\theta}$ を**推定可能関数**という。

正規方程式の解 $\hat{\boldsymbol{\theta}}$ は制約条件の与え方により異なる値が得られる。しかし推定可能関数について，$\boldsymbol{l}^T\hat{\boldsymbol{\theta}}$ は観測ベクトル $\boldsymbol{y}$ の線形結合として一意に決まる:

$$\begin{aligned}\boldsymbol{l}^T\hat{\boldsymbol{\theta}}\;&=\;\lambda_1y_1+\cdots+\lambda_Ny_N=\boldsymbol{\lambda}^T\,\boldsymbol{y}\\ &\qquad\mathrm{E}[\boldsymbol{\lambda}^T\,\boldsymbol{y}]=\boldsymbol{l}\,\boldsymbol{\theta},\quad \mathrm{V}[\boldsymbol{\lambda}^T\,\boldsymbol{y}]=\sigma^2\,\boldsymbol{\lambda}^T\boldsymbol{\lambda}\end{aligned} \tag{6}$$

自由度 1 の平方和

$$SS(\boldsymbol{l}^T\,\hat{\boldsymbol{\theta}})=(\boldsymbol{\lambda}^T\boldsymbol{y})^2/\boldsymbol{\lambda}^T\boldsymbol{\lambda} \tag{7}$$

推定可能関数の意味づけ

われわれが推測に利用できるのは観測データ $\boldsymbol{y}$ である。このとき，線形モデルの定義式 $\boldsymbol{y}=X\boldsymbol{\theta}+\boldsymbol{e}$ から分かるように，パラメータ $\boldsymbol{\theta}$ からは常に X の行ベクトル（すなわち，X^T の列ベクトル）を通して情報が観測データ $\boldsymbol{y}$ に伝えられている。したがって，線形モデル $\boldsymbol{y}=X\boldsymbol{\theta}+\boldsymbol{e}$ のパラメータ $\boldsymbol{\theta}$ に関する推定において，本質的に意味をもつのは X の横ベクトルにより表わされる推定可能関数 $\boldsymbol{l}^T\boldsymbol{\theta}$ である。

2.3 推定可能関数の例（アンバランストな一元配置）

アンバランストな一元配置モデル（2 水準; $n_1=3$, $n_2=2$）を考える。

$$\begin{pmatrix}y_1\\y_2\\y_3\\y_4\\y_5\end{pmatrix}=\begin{array}{c}\begin{array}{ccc}\mu&\alpha_1&\alpha_2\end{array}\\ \begin{pmatrix}1&1&0\\1&1&0\\1&1&0\\1&0&1\\1&0&1\end{pmatrix}\end{array}\begin{pmatrix}\mu\\\alpha_1\\\alpha_2\end{pmatrix}+\begin{pmatrix}e_1\\e_2\\e_3\\e_4\\e_5\end{pmatrix}$$

$\mathrm{rank}(X)=2<3=p$ であるから，μ, α_1, α_2 そのものは推定可能ではない。X の横ベクトルの線形結合として表わされる推定可能関数は,

$$\mu+\alpha_1,\quad \mu+\alpha_2,\quad \alpha_1-\alpha_2$$

などである。

このモデルでは制約条件の与え方によって，異なる正規方程式の解が得られる。しかし，どの解を用いても推定可能関数は観測ベクトル $\boldsymbol{y}$ の線形結合として一意的に表わされる（表 1)。

表 1　制約条件と推定可能関数（アンバランストな一元配置）

制約条件	$\hat{\alpha}_2 = 0$	$3\hat{\alpha}_1 + 2\hat{\alpha}_2 = 0$	$\hat{\alpha}_1 + \hat{\alpha}_2 = 0$
正規方程式の解	$\hat{\mu} = \frac{y_4+y_5}{2}$ $\hat{\alpha}_1 = \frac{y_1+y_2+y_3}{3} - \frac{y_4+y_5}{2}$ $\hat{\alpha}_2 = 0$	$\hat{\mu} = \frac{y_1+y_2+y_3+y_4+y_5}{5}$ $\hat{\alpha}_1 = \frac{2y_1+2y_2+2y_3-3y_4-3y_5}{15}$ $\hat{\alpha}_2 = \frac{-2y_1-2y_2-2y_3+3y_4+3y_5}{10}$	$\hat{\mu} = \frac{y_1+y_2+y_3}{6} + \frac{y_4+y_5}{4}$ $\hat{\alpha}_1 = \frac{y_1+y_2+y_3}{6} - \frac{y_4+y_5}{4}$ $\hat{\alpha}_2 = -\frac{y_1+y_2+y_3}{6} + \frac{y_4+y_5}{4}$
推定可能関数	$\hat{\mu} + \hat{\alpha}_1 = \frac{y_1+y_2+y_3}{3}$ $\hat{\mu} + \hat{\alpha}_2 = \frac{y_4+y_5}{2}$ $\hat{\alpha}_1 - \hat{\alpha}_2 = \frac{y_1+y_2+y_3}{3} - \frac{y_4+y_5}{2}$		

3　アンバランストな二元配置

3.1　数値例と線形モデル

それぞれ 2 水準の因子 A と B を考える。A_1B_1 セルだけ 2 回の繰返しがある。

表 2　2×2 データ

	B_1	B_2
A_1	y_1, y_2	y_3
A_2	y_4	y_5

表 3　2×2 数値例

	B_1	B_2
A_1	4.2, 3.8	1.0
A_2	0.0	0.0

線形モデルによる表現

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} = \begin{array}{c} \begin{array}{ccccccccc} \mu & \alpha_1 & \alpha_2 & \beta_1 & \beta_2 & (\alpha\beta)_{11} & (\alpha\beta)_{12} & (\alpha\beta)_{21} & (\alpha\beta)_{22} \end{array} \\ \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \boldsymbol{\theta} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{pmatrix} \tag{8}$$

推定可能関数と推定量

$$\begin{aligned} \mu_{11} &= \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11}, \quad \hat{\mu}_{11} = (y_1 + y_2)/2 = 4.0 \\ \mu_{12} &= \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12}, \quad \hat{\mu}_{12} = y_3 = 1.0 \\ \mu_{21} &= \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21}, \quad \hat{\mu}_{21} = y_4 = 0.0 \\ \mu_{22} &= \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22}, \quad \hat{\mu}_{22} = y_5 = 0.0 \end{aligned}$$

因子 A の主効果

- $\alpha_1 - \alpha_2 \Longrightarrow$ 推定可能ではない（X の横ベクトルをどのように組み合わせても，この形は得られない)。
- $\alpha_1 - \alpha_2 + \boxed{(\alpha\beta)_{ij} \text{ の組合せ}}$（必ず交互作用の項 $(\alpha\beta)_{ij}$ が入る）

  ↓

  TypeII, TypeIII (TypeIV) の違い

  因子 B の主効果についても同様である。

交互作用効果

- $(\alpha\beta)_{11} - (\alpha\beta)_{12} - (\alpha\beta)_{21} + (\alpha\beta)_{22} \Longrightarrow$ 推定可能（推定量は一意に決まる)。
  TypeII, III, IV の平方和は同じになる。

3.2　4 つの平方和

Proc GLM による計算プログラム（出力 4）と計算された 4 つの平方和（表 5）は以下のとおりである。

出力 4　アンバランスト二元配置データの計算

```
Data unbalanced2x2;
  Input A B y;
  DataLines;
    1 1 4.2
    1 1 3.8
    1 2 1.0
    2 1 0.0
    2 2 0.0
    ;
Run;
Proc GLM Data=unbalanced2x2;
  Class A B;
  Model y =  A B A*B / SS1 SS2 SS3 SS4;
Run;
```

表 5　アンバランスト二元配置データの 4 つの平方和

変動因	TypeI SS	TypeII SS	TypeIII SS	TypeIV SS
A	10.800	8.595	7.143	7.143
B	3.429	3.429	2.571	2.571
$A\times B$	2.571	2.571	2.571	2.571
誤差	0.080	0.080	0.080	0.080

交互作用 $A\times B$ は推定可能なので，どのタイプの平方和も値は同じになる。

3.3　主効果の TypeII と TypeIII 推定可能関数

因子 A の主効果（自由度 1）の推定を考える。TypeII，TypeIII（欠測セルが無いときは TypeIII=TypeIV）のいずれにおいても，まず主効果 B の影響を取り除く。そのためには，因子 B の各水準において因子 A の効果を計算する。すなわち，A の効果に関して次の 2 つの推定値が得られる（図 1）：

$$\begin{aligned}
\hat{y}_{A1} &= \hat{\mu}_{11} - \hat{\mu}_{21} = \frac{y_1 + y_2}{2} - y_4 = 4.0 \quad (B_1 \text{ 水準における } A \text{ の効果}) \\
&\quad \mathrm{V}[\hat{y}_{A1}] = \left(\frac{1}{2} + 1\right)\sigma^2 = \frac{3}{2}\sigma^2 \\
\hat{y}_{A2} &= \hat{\mu}_{12} - \hat{\mu}_{22} = y_3 - y_5 = 1.0 \quad (B_2 \text{ 水準における } A \text{ の効果}) \\
&\quad \mathrm{V}[\hat{y}_{A2}] = 2\,\sigma^2
\end{aligned}$$

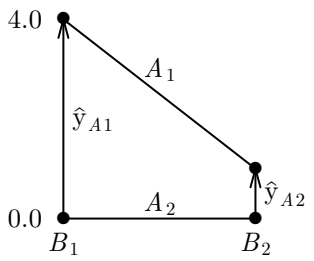

図 1　B の各水準での A の効果

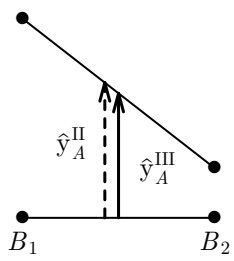

図 2　TypeII, III による A の効果の推定

A の主効果（自由度 1）の推定

$$\hat{y}_A = c_1\,\hat{y}_{A1} + c_2\,\hat{y}_{A2}$$

TypeII による推定: $c_1 \propto 1/\mathrm{V}[\hat{y}_{A1}], \quad c_2 \propto 1/\mathrm{V}[\hat{y}_{A2}]$

$$\hat{y}_A^{\mathrm{II}} = \frac{4}{7}\hat{y}_{A1} + \frac{3}{7}\hat{y}_{A2} = \frac{2y_1 + 2y_2 + 3y_3 - 4y_4 - 3y_5}{7} = \frac{19}{7}$$

$$\mathrm{E}[\hat{y}_A^{\mathrm{II}}] = \alpha_1 - \alpha_2 + \frac{4(\alpha\beta)_{11} + 3(\alpha\beta)_{12} - 4(\alpha\beta)_{21} - 3(\alpha\beta)_{22}}{7}$$

$$\mathrm{V}[\hat{y}_A^{\mathrm{II}}] = \frac{4+4+9+16+9}{49}\sigma^2 = \frac{6}{7}\sigma^2$$

$$\text{平方和:} \quad SS(\hat{y}_A^{\mathrm{II}}) = \left(\frac{19}{7}\right)^2 \Big/ \frac{6}{7} = \frac{19^2}{42} = 8.595$$

TypeIII による推定: $c_1 = c_2 = 1/2$（サンプルサイズに関係なく全てのセルを同等に扱う）

$$\hat{y}_A^{\mathrm{III}} = \frac{1}{2}\hat{y}_{A1} + \frac{1}{2}\hat{y}_{A2} = \frac{y_1 + y_2 + 2y_3 - 2y_4 - 2y_5}{4} = \frac{5}{2}$$

$$\mathrm{E}[\hat{y}_A^{\mathrm{III}}] = \alpha_1 - \alpha_2 + \frac{(\alpha\beta)_{11} + (\alpha\beta)_{12} - (\alpha\beta)_{21} - (\alpha\beta)_{22}}{2}$$

$$\mathrm{V}[\hat{y}_A^{\mathrm{III}}] = \frac{1+1+4+4+4}{16}\sigma^2 = \frac{7}{8}\sigma^2$$

$$\text{平方和:} \quad SS(\hat{y}_A^{\mathrm{III}}) = \left(\frac{5}{2}\right)^2 \Big/ \frac{7}{8} = \frac{50}{7} = 7.143$$

図 2 の TypeII と TypeIII の主効果の推定値を示す。TypeIII では B_1 水準と B_2 水準の中央で A の主効果を推定しているのに対し，TypeII ではサンプルサイズの大きい方に移動している。

3.4 TypeII による推定の別の解釈

主効果に関する TypeII の平方和は，モデルステートメントで

(1) `Model y = A B A*B;`

(2) `Model y = A B;`

のどちらを指定しても同じになる（なお (2) のモデルでは，$\alpha_1 - \alpha_2$ は推定可能となり，推定量は一意に決まる）。

(1) のようにモデルステートメントに `A*B` を含めるということは，われわれは交互作用を考慮して解析を進めたいということを `GLM` に伝えているのである。このとき TypeII による推定では，モデル (2) を仮定して計算を進めている。すなわち，TypeII による推定では，あたかも交互作用が存在しないかのようにみなして（つまり (2) のモデルを仮定して）主効果を推定していることになる。

3.5 アンバランスの程度が大きいときの比較

セル間でサンプルサイズが大きく異なる場合の例を図 3 と図 4 に示す。TypeII ではサンプルサイズの大きい側に引かれる。図 4 のようにサンプルサイズの大きい側で因子 A の効果が小さければ，因子 A の主効果の TypeII による推定値は小さくなり，平方和も小さくなる。したがって，一般的には TypeIII よりも TypeII の方が主効果の検出力が高いとはいえない。すなわち，主効果に関して

- TypeII で有意で，TypeIII で有意ではない。
- TypeII で有意ではなく，TypeIII で有意。

のどちらの場合も起こりうる。

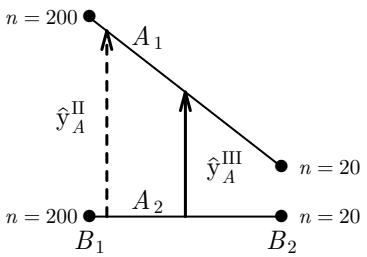

図 3　大きいサンプルサイズ側で大きな効果

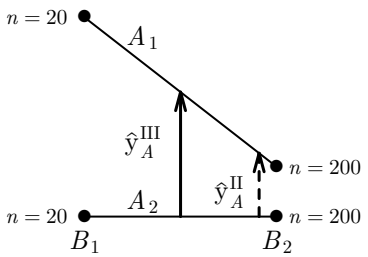

図 4　大きいサンプルサイズ側で小さな効果

実験計画に基づいたデータであればサンプルサイズが大きく異なることは少ない。しかし調査研究においては，処理組合せによってサンプルサイズの大きなアンバランスが生じることがある。そのとき，サンプルサイズの大きさがマーケットサイズ（期待される利益）を表わしているのであれば，そのサンプルサイズに応じて効果を評価する TypeII は意味がある。一方，大きなサンプルサイズが，単にデータが集まりやすかったということの結果である場合は注意が必要である。たとえば次のような場合である。

- 医薬データ

  因子 A: 薬剤の種類（A_1: 新薬剤，A_2: 現行薬剤）

  因子 B: 調査対象者の生活習慣（B_1: 特異（あるいは特定の疾患あり），B_2: 正常）
- 農業データ

  因子 A: 品種（A_1: 新品種，A_2: 対照品種）

  因子 B: 耕種方法（B_1: 有機栽培，B_2: 慣行栽培）

一方，TypeIII による効果の推定の考え方は，各セルの標本平均 $\hat{\mu}_{ij}$ は交互作用を含んだ母平均 μ_{ij} に関する情報を含んでいるのだから，サンプルサイズに関係なく各セルの標本平均を平等に扱おうというものである。ただし TypeIII による推定では，サンプルサイズの小さいセルも同等に扱うため，極端にサンプルサイズの小さいセルが存在すると結果が不安定になる。

3.6 アンバランストなデータを解析するときの対策

アンバランストな多元配置データを解析するときは，`Model` ステートメントに `SS2` と `SS3` の両方を指定するとよい。デフォルト（平方和のタイプを指定しない），あるいは `SS2` のみを指定すると，一方の平方和しか出力されないので注意が必要である。

```
Model y = A B A*B;            ⟶  TypeI と TypeIII が出力される。
Model y = A B A*B / SS2;      ⟶  TypeII のみが出力される。
Model y = A B A*B / SS2 SS3;  ⟶  TypeII と TypeIII が出力される。
```

(1) 交互作用が有意

想定される出力結果を表 6 に示す。まず交互作用 $A{\times}B$ に関しては，推定可能であるから，TypeII の結果と TypeIII の結果は同じになる（どちらも有意，あるいはどちらも有意ではない）。交互作用が有意ということは，因子 A の効果が他方の因子 B の水準ごとに異なるということであるから，主効果の検定はあまり重要ではない。セル平均の $A{\times}B$ 二元表や図 1 ～ 4 のような応答図によって，どのような形で交互作用が生じているのかを検討することに意味がある。

(2) 交互作用が有意ではないとき

交互作用が有意でなくても，その影響により TypeII と TypeIII とで主効果に対する検定結果が異なることが起こりうる。まず主効果の検定において，TypeII と TypeIII の結果が同じ（どちらも有意，あるいはどちらも有意ではない）であればその結果に従えばよい。

主効果に関して，TypeII と TypeIII の結果が異なる場合は，サンプルサイズのアンバランスが影響していることになる。各セルの標本平均とサンプルサイズの表から，第 3.3 節に述べた方法により，なぜ異なる結果になったのかを検討することが有効である。このとき，平方和だけを見たのでは，どこに原因があるのかを探索することは困難である。

表 6　アンバランストモデルの分散分析の解釈

変動因	TypeII	TypeIII	TypeII		TypeIII	TypeII		TypeIII
A			F 検定（p 値）	$=$	F 検定（p 値）	F 検定（p 値）	$\neq$	F 検定（p 値）
B								
$A{\times}B$	有意	有意	NS ($p = 0.2$)		NS ($p = 0.2$)	NS ($p = 0.2$)		NS ($p = 0.2$)
	交互作用が有意		交互作用が有意ではない			交互作用が有意ではない		

4　直交表実験における欠測値

直交表実験は，高度にバランスが保たれていることを利用して効率的に情報を獲得するための実験計画手法であり，とくに我が国において，工業の品質管理分野や農業実験において技術開発に大きく貢献した。処理組合せに欠測が生じると高度なバランスが保たれなくなるため，従来はデータの解析が困難であった。しかし，GLM プロシージャを利用することにより，アンバランストな多元配置実験として解析することができる。

表 7　欠測のある L_8 直交表実験

列	(1)	(2)	(3)	(4)	(5)	
因子	A	B	$A\times B$	C	$A\times C$	
No.					D	y
1	1	1	1	1	1	9.7
2	1	1	1	2	2	6.6
3	1	2	2	1	1	8.4
4	1	2	2	2	2	6.4
5	2	1	2	1	2	2.5
6	2	1	2	2	1	1.4
7	2	2	1	1	2	4.9
8	2	2	1	2	1	欠測

ここでは，簡単な例として実験単位の数が 8 の $L_8(2^7)$ 直交表（表 7）を考える。第 (5) 列は交互作用 $A\times C$ を表わしている。しかし，この交互作用は小さいと考えられたので，この列に別の因子 D を割り付けた（要因 $A\times C$ と D が交絡しているという）。No.8 の処理組合せで欠測が生じた。計算プログラムを出力 8 に示す。

出力 8　第 (5) 列を交互作用 $A\times C$，あるいは主効果 D として分散分析

```
Data orthdat;
  Input A B C y;
  D = MOD(A+C, 2) + 1;  /* D=AxC */
  Output;
  DataLines;
    1 1 1 9.7
    1 1 2 6.6
    1 2 1 8.4
    1 2 2 6.4
    2 1 1 2.5
    2 1 2 1.4
    2 2 1 4.9
    ;
Run;

/* Column (5) = AxC */
Proc GLM Data=orthdat;
  Class A B C;
  Model y =  A B C A*B A*C / SS1 SS2 SS3 SS4;
Run;

/* Column (5) = D */
Proc GLM Data=orthdat;
  Class A B C D;
  Model y =  A B C A*B D / SS1 SS2 SS3 SS4;
Run;
```

表 9 に，第 (5) 列を交互作用 $A\times C$ と見なして計算したときの 4 つの平方和を示す。

表 9　第 (5) 列を交互作用 $A{\times}C$ として計算

変動因	TypeI SS	TypeII SS	TypeIII SS	TypeIV SS
A	40.186	43.264	28.521	28.521
B	0.888	0.135	0.908	0.908
C	9.216	6.407	4.441	4.441
$A{\times}B$	2.667	3.308	3.308	3.308
$A{\times}C$	0.701	0.701	0.701	0.701

注 表 9 において，因子 C の主効果に関して，TypeI と TypeII の平方和が異なることに注意。因子 C の主効果の TypeII の平方和は

$$ModelSS(A,\ B,\ C,\ A{\times}B) - ModelSS(A,\ B,\ A{\times}B)$$

により計算される。すなわち，TypeII の平方和は最初に主効果のみをモデルに取り込むわけではない。

一方，表 10 に，第 (5) 列を主効果 D と見なして計算したときの 4 つの平方和を示す。

表 10　第 (5) 列を主効果 D として計算

変動因	TypeI SS	TypeII SS	TypeIII SS	TypeIV SS
A	40.186	39.784	28.521	28.521
B	0.888	0.135	0.908	0.908
C	9.216	4.441	4.441	4.441
$A{\times}B$	2.667	3.308	3.308	3.308
D	0.701	0.701	0.701	0.701

交絡している第 (5) 列を交互作用 $A{\times}C$ とみなすか（表 9），主効果 D とみなすか（表 10）によって，TypeII の平方和は値が異なる。したがって，直交表実験で欠測処理が生じた場合には，TypeII の平方和については注意が必要である。

一方 TypeIII の平方和は，第 (5) 列を交互作用 $A{\times}B$ として扱っても，あるいは主効果 D として扱っても値は同じになる。これは，よく行なわれる説明

- TypeIII の計算では，交互作用も 1 つの要因として扱う

に対応するものである。

5　TypeIV の推定

5.1　欠測セルのある二元配置

欠測セルが無ければ TypeIII の推定と TypeIV の推定は同じものになる。表 3 に対して，因子 B に水準を 1 つ追加したものを表 11 に示す。ただし A_1B_3 セルは欠測とする。また表 12 は因子 B の水準 1 と 3 を交換したものである。

表 11　欠測セルのある 2×3 数値例

	B_1	B_2	B_3
A_1	4.2, 3.8	1.0	×
A_2	0.0	0.0	0.0

表 12　B_1 と B_3 を交換

	B_1	B_2	B_3
A_1	×	1.0	4.2, 3.8
A_2	0.0	0.0	0.0

それぞれに対して `GLM` で計算した 4 つの平方和を表 13 と表 14 に示す。TypeIV 平方和は，水準番号を付けかえると異なる値が得られる。そして

```
NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.
```

というコメントが出力される。すなわち，TypeIV の平方和の計算は複数の仮説が考えられている。

表 13　第 11 の分散分析

変動因	TypeI SS	TypeII SS	TypeIII SS	TypeIV SS
A	13.500	8.595	7.143	7.143
B	3.429	3.429	2.571	0.000
$A\times B$	2.571	2.571	2.571	2.571

TypeIV SS のコメント出力
NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

表 14　第 12 の分散分析

変動因	TypeI SS	TypeII SS	TypeIII SS	TypeIV SS
A	13.500	8.595	7.143	7.143
B	3.429	3.429	2.571	3.000
$A\times B$	2.571	2.571	2.571	2.571

TypeIV SS のコメント出力
NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

5.2　TypeIV の推定可能関数

水準数が 3 以上（自由度が 2 以上）の場合，自由度 1 の成分に分解して推定が行なわれる。水準数が 3 の場合

(1) B_1 vs. B_3:　$\beta_1 - \beta_3 + \boxed{\text{交互作用項 } (\alpha\beta)_{ij}}$

(2) B_2 vs. B_3:　$\beta_2 - \beta_3 + \boxed{\text{交互作用項 } (\alpha\beta)_{ij}}$

の 2 つの成分に分解して効果が推定される。さらに TypeIV の計算においては，データの揃っているセルを用いて効果を推定する。

表 11 の TypeIV 平方和の計算の様子を図 5 と図 6 に示す。まず B_1 vs. B_3 の成分の計算において，A_1B_3 セルは欠測なので，A_2B_1 セルと A_2B_3 セルのデータのみが使われる（図 5）。同様に，B_2 vs. B_3 の成分の計算においても，A_1B_3 セルが欠測なので，A_2B_2 セルと A_2B_3 セルのデータのみが使われる（図 6）。したがって，表 11 のデータセットに対する主効果 B の TypeIV 平方和はゼロになる。

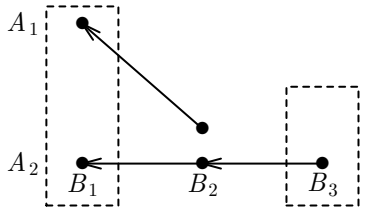

図 5　成分 (1) の推定

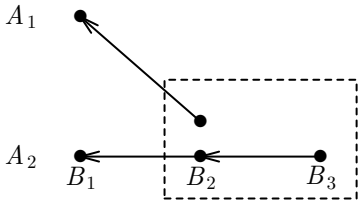

図 6　成分 (2) の推定

次に水準 B_1 と水準 B_3 を入れかえたとき（表 12）の TypeIV 平方和の計算の様子を図 7 と図 8 に示す。B_1 vs. B_3 の成分の計算では, A_1B_1 セルは欠測なので, A_2B_1 セルと A_2B_3 セルのデータのみが使われる（図 7）。次に B_2 vs. B_3 の成分を考えると，4 つのセルが欠測なく揃っている（図 8）。そこで，この 4 つのセルのデータを使って TypeIII 方式で推定可能関数を計算し，さらに平方和が計算される。したがって，表 12 のデータセットに対する主効果 B の TypeIV 平方和はゼロではなくなる。

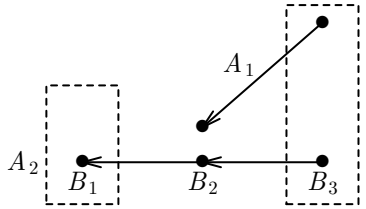

図 7　水準の交換後の成分 (1) の推定

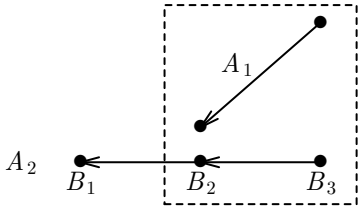

図 8　水準の交換後の成分 (2) の推定

6　おわりに

本講演では，アンバランストな多元配置データの解析に対して，推定可能関数の観点から TypeII, TypeIII, TypeIV の違いを考えた。欠測セルが無い場合は，TypeIII と TypeIV の解析の結果は同じになるので，TypeII を使うか，TypeIII を使うかの選択になる。対策としては第 3.6 節に述べたように，TypeII と TypeIII の結果の両方を出力して，その違いを検討することが有効である。そのとき，線形モデルの推定可能関数の考え方が役に立つ。

欠測セルがある場合は，TypeII, TypeIII, TypeIV で異なった結果が得られる。第 5.2 節で説明したように，TypeIV の解析では水準番号の付けかえによって異なる結果が得られる。しかし水準番号の付けかえはデータ解析において本質的な事柄ではないので，TypeIV を使って解析することは勧められない。欠測セルがある場合も，TypeII か TypeIII で解析すればよい。TypeII を使うか TypeIII を使うかの議論は，欠測セルの無い場合と同様に，各セルのサンプルサイズを考慮して解析する (TypeII) か，各セルをサンプルサイズに関係なく同等に扱う (TypeIII) かを判断して決めればよい。

参考文献

[1] 三輪哲久: 実験計画法と分散分析, 朝倉書店, 2015.

[2] SAS Institute Inc.: *SAS/STAT© 9.3 User's Guide*, SAS Institute Inc., 2011.

[3] 高橋行雄, 大橋靖雄, 芳賀敏郎: SAS による実験データの解析, 東京大学出版会, 1989.